Complete Nucleotide and Deduced Amino Acid Sequence of Rice homolog of MLH1

1 CGGCACGAGATTTTGCAGTCTCCTCTCCTCCTCCGCTCGAGCGAGTGAGTCCCGACCACG 60

61 TCGCTGCCCTCGCCTCACCGCCGGCCAACCGCCGTGACGAGAGATCGAGCAGGGCGGGGC 120

121 ATGGACGAGCCTTCGCCGCGCGGAGGTGGGTGCGCCGGGGAGCCGCCCCGCATCCGGAGG 180
MetAspGluProSerProArgGlyGlyGlyCysAlaGlyGluProProArgIleArgArg

181 TTGGAGGAGTCGGTGGTGAACCGCATCGCGGCGGGGGAGGTGATCCAGCGGCCGTCGTCG 240
LeuGluGluSerValValAsnArgIleAlaAlaGlyGluValIleGlnArgProSerSer

241 GCGGTGAAGGAGCTCATCGAGAACAGCCTCGACGCTGGCGCCTCCAGCGTCTCCGTTGCG 300
AlaValLysGluLeuIleGluAsnSerLeuAspAlaGlyAlaSerSerValSerValAla

301 GTGAAGGACGGTGGCCTCAAGCTCATCCAGGTCTCCGATGACGGCCATGGCATCAGGTTT 360
ValLysAspGlyGlyLeuLysLeuIleGlnValSerAspAspGlyHisGlyIleArgPhe

361 GAGGATTTGGCAATATTGTGCGAAAGGCATACTACCTCAAAGTTATCTGCATACGAGGAT 420
GluAspLeuAlaIleLeuCysGluArgHisThrThrSerLysLeuSerAlaTyrGluAsp

421 CTGCAGACCATAAAATCGATGGGGTTCAGAGGGGAGGCTTTGGCTAGTATGACTTATGTT 480
LeuGlnThrIleLysSerMetGlyPheArgGlyGluAlaLeuAlaSerMetThrTyrVal

481 GGCCATGTTACCGTGACAACGATAACAGAAGGCCAATTGCACGGCTACAGGGTTTCTTAC 540
GlyHisValThrValThrThrIleThrGluGlyGlnLeuHisGlyTyrArgValSerTyr

541 AGAGATGGTGTAATGGAGAATGAGCCTAAGCCTTGCGCTGCGGTGAAAGGAACTCAAGTC 600
ArgAspGlyValMetGluAsnGluProLysProCysAlaAlaValLysGlyThrGlnVal

601 ATGGTTGAAAATCTATTTTACAACATGGTAGCCCGCAAGAAAACATTGCAGAACTCCAAT 660
MetValGluAsnLeuPheTyrAsnMetValAlaArgLysLysThrLeuGlnAsnSerAsn

661 GATGACTACCCCAAGATCGTAGACTTCATCAGTCGGTTTGCAGTCCATCACATCAACGTT 720
AspAspTyrProLysIleValAspPheIleSerArgPheAlaValHisHisIleAsnVal

721 ACCTTCTCTTGCAGAAAGCATGGAGCCAATAGAGCAGATGTTCATAGTGCAAGTACATCC 780
ThrPheSerCysArgLysHisGlyAlaAsnArgAlaAspValHisSerAlaSerThrSer

781 TCAAGGTTAGATGCTATCAGGAGTGTCTATGGGGCTTCTGTCGTTCGTGATCTCATAGAA 840
SerArgLeuAspAlaIleArgSerValTyrGlyAlaSerValValArgAspLeuIleGlu

FIGURE 1A

```
841   ATAAAGGTTTCATATGAGGATGCTGCAGATTCAATCTTCAAGATGGATGGTTACATCTCA   900
      IleLysValSerTyrGluAspAlaAlaAspSerIlePheLysMetAspGlyTyrIleSer

901   AATGCAAATTATGTGGCAAAGAAGATTACAATGATTCTTTTCATAAATGATAGGCTTGTA   960
      AsnAlaAsnTyrValAlaLysLysIleThrMetIleLeuPheIleAsnAspArgLeuVal

961   GACTGTACTGCTTTGAAAAGAGCTATTGAATTTGTGTACTCTGCAACATTGCCTCAAGCA   1020
      AspCysThrAlaLeuLysArgAlaIleGluPheValTyrSerAlaThrLeuProGlnAla

1021  TCCAAACCTTTCATATACATGTCCATACATCTTCCATCAGAACACGTGGATGTTAATATA   1080
      SerLysProPheIleTyrMetSerIleHisLeuProSerGluHisValAspValAsnIle

1081  CACCCAACCAAGAAAGAGGTTAGCCTTTTGAATCAAGAGCGTATTATTGAAACAATAAGA   1140
      HisProThrLysLysGluValSerLeuLeuAsnGlnGluArgIleIleGluThrIleArg

1141  AATGCTATTGAGGAAAAACTGATGAATTCTAATACAACCAGGATATTCCAAACTCAGGCA   1200
      AsnAlaIleGluGluLysLeuMetAsnSerAsnThrThrArgIlePheGlnThrGlnAla

1201  TTAAACTTATCAGGGATTGCTCAAGCTAACCCACAAAAGGATAAGGTTTCTGAGGCCAGT   1260
      LeuAsnLeuSerGlyIleAlaGlnAlaAsnProGlnLysAspLysValSerGluAlaSer

1261  ATGGGTTCTGGAACAAAATCTCAAAAAATTCCTGTGAGCCAAATGGTCAGAACAGATCCA   1320
      MetGlySerGlyThrLysSerGlnLysIleProValSerGlnMetValArgThrAspPro

1321  CGCAATCCATCTGGAAGATTGCACACCTACTGGCACGGGCAATCTTCAAATCTTGAAAAG   1380
      ArgAsnProSerGlyArgLeuHisThrTyrTrpHisGlyGlnSerSerAsnLeuGluLys

1381  AAATTTGATCTTGTATCTGTAAGAAATGTTGTAAGATCAAGGAGAAACCAAAAAGATGCT   1440
      LysPheAspLeuValSerValArgAsnValValArgSerArgArgAsnGlnLysAspAla

1441  GGTGATTTGTCAAGCCGTCATGAGCTCCTTGTGGAAATAGATTCTAGCTTCCATCCTGGC   1500
      GlyAspLeuSerSerArgHisGluLeuLeuValGluIleAspSerSerPheHisProGly

1501  CTTTTGGACATTGTCAAGAACTGCACATATGTTGGACTTGCCGATGAAGCCTTTGCTTTG   1560
      LeuLeuAspIleValLysAsnCysThrTyrValGlyLeuAlaAspGluAlaPheAlaLeu

1561  ATACAACACAATACCCGCTTATACCTTGTAAATGTGGTAAATATTAGTAAAGAACTTATG   1620
      IleGlnHisAsnThrArgLeuTyrLeuValAsnValValAsnIleSerLysGluLeuMet

1621  TACCAGCAAGCTTTGTGCCGTTTTGGGAACTTCAATGCTATTCAGCTCAGTGAACCAGCT   1680
      TyrGlnGlnAlaLeuCysArgPheGlyAsnPheAsnAlaIleGlnLeuSerGluProAla
```

FIGURE 1B

1681 CCACTTCAGGAGTTGCTGGTGATGGCACTGAAAGACGATGAATTGATGAGTGATGAAAAG 1740
ProLeuGlnGluLeuLeuValMetAlaLeuLysAspAspGluLeuMetSerAspGluLys

1741 GATGATGAGAAACTGGAGATTGCAGAAGTAAACACTGAGATACTAAAAGAAAATGCTGAG 1800
AspAspGluLysLeuGluIleAlaGluValAsnThrGluIleLeuLysGluAsnAlaGlu

1801 ATGATTAATGAGTACTTTTCTATTCACATTGATCAAGATGGCAAATTGACAAGACTTCCT 1860
MetIleAsnGluTyrPheSerIleHisIleAspGlnAspGlyLysLeuThrArgLeuPro

1861 GTTGTACTGGACCAGTACACCCCTGATATGGACCGTCTTCCAGAATTTGTGTTGGCTTTA 1920
ValValLeuAspGlnTyrThrProAspMetAspArgLeuProGluPheValLeuAlaLeu

1921 GGAAATGATGTTACTTGGGATGACGAGAAAGAGTGCTTCAGAACAGTAGCTTCTGCTGTA 1980
GlyAsnAspValThrTrpAspAspGluLysGluCysPheArgThrValAlaSerAlaVal

1981 GGAAACTTCTATGCACTTCATCCCCCAATCCTTCCAAATCCATCTGGGAATGGCATTCAT 2040
GlyAsnPheTyrAlaLeuHisProProIleLeuProAsnProSerGlyAsnGlyIleHis

2041 TTATACAAGAAAAATAGAGATTCAATGGCTGATGAACATGCTGAGAATGATCTAATATCA 2100
LeuTyrLysLysAsnArgAspSerMetAlaAspGluHisAlaGluAsnAspLeuIleSer

2101 GATGAAAATGACGTTGATCAAGAACTTCTTGCGGAAGCAGAAGCAGCATGGGCCCAACGT 2160
AspGluAsnAspValAspGlnGluLeuLeuAlaGluAlaGluAlaAlaTrpAlaGlnArg

2161 GAGTGGACCATTCAGCATGTCTTGTTTCCATCCATGCGACTTTTCCTCAAGCCCCCGAAG 2220
GluTrpThrIleGlnHisValLeuPheProSerMetArgLeuPheLeuLysProProLys

2221 TCAATGGCAACAGATGGAACGTTTGTGCAGGTTGCTTCCTTGGAGAAACTCTACAAGATT 2280
SerMetAlaThrAspGlyThrPheValGlnValAlaSerLeuGluLysLeuTyrLysIle

2281 TTTGAAAGGTGTTAGCTCATAAGTGAGAAAATGAAGGCAGAGTAAGATCATGATTCATGG 2340
PheGluArgCysEnd

2341 AGTGTTTTTGAAAATGTGTATAATTTCACCGTATTATGTACTTTGATAGTGTCTGTAGAA 2400

2401 ACTGAAGAAAGAAAGATGGCTTTACTTCTGAATTGAAAGTTAACGATGCCAGCAATTGTA 2460

2461 TATTCTGATCAACCAAAAAAAAAAAAAAAAAAAAAAAAAAA 2501

FIGURE 1C

<u>AminoAcid Sequence of Rice Homolog[illegible] of MLH1.</u>

1	MDEPSPRGGG	CAGEPPRIRR	LEESVVNRIA	AGEVIQRPSS	AVKELIENSL
51	DAGASSVSVA	VKDGGLKLIQ	VSDDGHGIRF	EDLAILCERH	TTSKLSAYED
101	LQTIKSM**GFR**	**GEALASMTY**V	GHVTVTTITE	GQLHGYRVSY	RDGVMENEPK
151	PCAAVKGTQV	MVENLFYNMV	ARKKTLQNSN	DDYPKIVDFI	SRFAVHHINV
201	TFSCRKHGAN	RADVHSASTS	SRLDAIRSVY	GASVVRDLIE	IKVSYEDAAD
251	SIFKMDGYIS	NANYVAKKIT	MILFINDRLV	DCTALKRAIE	FVYSATLPQA
301	SKPFIYMSIH	LPSEHVDVNI	HPTKKEVSLL	NQERIIETIR	NAIEEKLMNS
351	NTTRIFQTQA	LNLSGIAQAN	PQKDKVSEAS	MGSGTKSQKI	PVSQMVRTDP
401	RNPSGRLHTY	WHGQSSNLEK	KFDLVSVRNV	VRSRRNQKDA	GDLSSRHELL
451	VEIDSSFHPG	LLDIVKNCTY	VGLADEAFAL	IQHNTRLYLV	NVVNISKELM
501	YQQALCRFGN	FNAIQLSEPA	PLQELLVMAL	KDDELMSDEK	DDEKLEIAEV
551	NTEILKENAE	MINEYFSIHI	DQDGKLTRLP	VVLDQYTPDM	DRLPEFVLAL
601	GNDVTWDDEK	ECFRTVASAV	GNFYALHPPI	LPNPSGNGIH	LYKKNRDSMA
651	DEHAENDLIS	DENDVDQELL	AEAEAAWAQR	EWTIQHVLFP	SMRLFLKPPK
701	SMATDGTFVQ	VASLEKLYKI	FERC*		

mutL/PMS1 signature sequence is shown in **bold**.

FIGURE 2

Amino Acid Sequence Comparison of Rice and Arabidopsis mutL Homologs

```
  2 DEPSPRGGGCAGEPPRIRRLEESVVNRIAAGEVIQRPSSAVKELIENSLD  51
    :| ||         |||:|.|||||||||||||||||||| ||||||:|||||
 13 EEESPATTIVPREPPKIQRLEESVVNRIAAGEVIQRPVSAVKELVENSLD  62

 52 AGASSVSVAVKDGGLKLIQVSDDGHGIRFEDLAILCERHTTSKLSAYEDL 101
    | .||:|| |||||||||||||||||||| ||| |||||||||||. :|||
 63 ADSSSISVVVKDGGLKLIQVSDDGHGIRREDLPILCERHTTSKLTKFEDL 112

102 QTIKSMGFRGEALASMTYVGHVTVTTITEGQLHGYRVSYRDGVMENEPKP 151
     .: ||||||||||||||| ||||||||:||:|||||||||||||.|||
113 FSLSSMGFRGEALASMTYVAHVTVTTITKGQIHGYRVSYRDGVMEHEPKA 162

152 CAAVKGTQVMVENLFYNMVARKKTLQNSNDDYPKIVDFISRFAVHHINVT 201
    |||||||:|||||||||:||:|||||| |||. |||| :|| |:|: ||.
163 CAAVKGTQIMVENLFYNMIARRKTLQNSADDYGKIVDLLSRMAIHYNNVS 212

202 FSCRKHGANRADVHSASTSSRLDAIRSVYGASVVRDLIEIKVSYEDAADS 251
    |||||||| :|||||   . ||||.|||||| || :.|..:.|| |..
213 FSCRKHGAVKADVHSVVSPSRLDSIRSVYGVSVAKNLMKVEVSSCDSSGC 262

252 IFKMDGYISNANYVAKKITMILFINDRLVDCTALKRAIEFVYSATLPQAS 301
    | |:|:|||.||||||  ::||||||||:|.|||||||| ||.||||.||
263 TFDMEGFISNSNYVAKKTILVLFINDRLVECSALKRAIEIVYAATLPKAS 312

302 KPFIYMSIHLPSEHVDVNIHPTKKEVSLLNQERIIETIRNAIEEKLMNSN 351
    |||:||||.|| ||||:||||||||||||||| ||| |.. :| || |.|
313 KPFVYMSINLPREHVDINIHPTKKEVSLLNQEIIIEMIQSEVEVKLRNAN 362

352 TTRIFQTQALNLSGIAQANPQKDKVSEASMGSGTKSQKIPVSQMVRTDPR 401
     || || | .   .    ||     .   || |.||:||..|||||
363 DTRTFQEQKVEYIQ.STLTSQKSDSPVSQKPSGQKTQKVPVNKMVRTDSS 411

402 NPSGRLHTYWHGQSSNLEKKFDLVS.VRNVVRSRRNQKDAGDLSSRHELL 450
    .|.|||| :    .  .| |   .| ||. || ||| |:  ||||  ||:
412 DPAGRLHAFLQPKPQSLPDKVSSLSVVRSSVRQRRNPKETADLSSVQELI 461

451 VEIDSSFHPGLLDIVKNCTYVGLADEAFALIQHNTRLYLVNVVNISKELM 500
     :||  |||:|: |:||||||:||: |||:|:|| ||| ||||:|||||
462 AGVDSCCHPGMLETVRNCTYVGMADDVFALVQYNTHLYLANVVNLSKELM 511

501 YQQALCRFGNFNAIQLSEPAPLQELLVMALKDDEL..MSDEKDDEKLEIA 548
    ||| | || .|||||||:|||| ||:.:|||:::|    .| ||| |  ||
512 YQQTLRRFAHFNAIQLSDPAPLSELILLALKEEDLDPGNDTKDDLKERIA 561

549 EVNTEILKENAEMINEYFSIHIDQDGKLTRLPVVLDQYTPDMDRLPEFVL 598
    |.|||:||| |||: ||||:|||     |.||||:|||||||||||.|||.|
562 EMNTELLKEKAEMLEEYFSVHIDSSANLSRLPVILDQYTPDMDRVPEFLL 611
```

FIGURE 3A

```
                .         .         .         .         .
599 ALGNDVTWDDEKECFRTVASAVGNFYALHPPILPNPSGNGIHLYKKNRDS 648
     ||||| |:||| ||. |..|:|||||:|||:||||||.||  | |  :|
612 CLGNDVEWEDEKSCFQGVSAAIGNFYAMHPPLLPNPSGDGIQFYSKRGES 661

                .         .         .         .         .
649 MADEHAENDLISDENDVDQELLAEAEAAWAQREWTIQHVLFPSMRLFLKP 698
     :.     :    |...||:||.:|| |||||||.|||||||||||||||
662 SQEKSDLEGNVDMEDNLDQDLLSDAENAWAQREWSIQHVLFPSMRLFLKP 711

                .         .
699 PKSMATDGTFVQVASLEKLYKIFERC 724
    | |||..||||.||||||||||||||
712 PASMASNGTFVKVASLEKLYKIFERC 737
```

Deduced amino acid sequences of *Oryza sativa* and *Arabidopsis thaliana* (Genbank ID, SP_PL:Q9ZRV4) were compared using the Bestfit program of GCG.

FIGURE 3B

Comparison of cDNA sequences of MLH1 orthologs from A. thaliana and O. sativa

```
158 GGGAGCCGCCCCGCATCCGGAGGTTGGAGGAGTCGGTGGTGAACCGCATC 207
 73 GAGAGCCACCGAAGATTCAACGCTTAGAAGAATCAGTAGTCAACCGTATC 122

208 GCGGCGGGGGAGGTGATCCAGCGGCCGTCGTCGGCGGTGAAGGAGCTCAT 257
123 GCAGCTGGTGAAGTAATCCAGCGTCCAGTTTCAGCTGTGAAAGAGCTCGT 172

258 CGAGAACAGCCTCGACGCTGGCGCCTCCAGCGTCTCCGTTGCGGTGAAGG 307
173 TGAGAACAGCCTCGACGCCGATTCAAGTTCCATAAGCGTCGTTGTCAAAG 222

308 ACGGTGGCCTCAAGCTCATCCAGGTCTCCGATGACGGCCATGGCATCAGG 357
223 ACGGTGGTTTGAAACTCATTCAAGTCTCCGACGACGGTCACGGTATTAGA 272

358 TTTGAGGATTTGGCAATATTGTGCGAAAGGCATACTACCTCAAAGTTATC 407
273 CGTGAAGACTTGCCGATACTATGCGAGAGACATACAACATCGAAGCTGAC 322

408 TGCATACGAGGATCTGCAGACCATAAAATCGATGGGGTTCAGAGGGGAGG 457
323 TAAGTTTGAGGATTTGTTCTCTCTGAGTTCAATGGGATTTAGAGGAGAGG 372

458 CTTTGGCTAGTATGACTTATGTTGGCCATGTTACCGTGACAACGATAACA 507
373 CATTAGCTAGTATGACCTATGTTGCTCATGTTACAGTGACTACTATTACT 422

508 GAAGGCCAATTGCACGGCTACAGGGTTTCTTACAGAGATGGTGTAATGGA 557
423 AAAGGCCAGATTCATGGTTATAGAGTGTCTTATAGAGATGGTGTCATGGA 472

558 GAATGAGCCTAAGCCTTGCGCTGCGGTGAAAGGAACTCAAGTCATGGTTG 607
473 GCATGAACCAAAGGCGTGTGCTGCTGTCAAAGGAACACAGATAATGGTGG 522

608 AAAATCTATTTTACAACATGGTAGCCCGCAAGAAAACATTGCAGAACTCC 657
523 AGAATTTGTTCTACAATATGATTGCTAGAAGGAAGACACTTCAAAATTCT 572

658 AATGATGACTACCCCAAGATCGTAGACTTCATCAGTCGGTTTGCAGTCCA 707
573 GCTGATGATTACGGGAAAATCGTGGATTTGCTGAGCCGGATGGCTATTCA 622

708 TCACATCAACGTTACCTTCTCTTGCAGAAAGCATGGAGCCAATAGAGCAG 757
623 TTACAATAATGTCAGCTTTTCTTGTCGAAAGCATGGAGCTGTTAAGGCTG 672

758 ATGTTCATAGTGCAAGTACATCCTCAAGGTTAGATGCTATCAGGAGTGTC 807
673 ATGTTCACTCAGTCGTGTCACCTTCAAGGCTTGATTCAATTAGGTCTGTA 722
```

FIGURE 4A

```
 808 TATGGGGCTTCTGTCGTTCGTGATCTCATAGAAATAAAGGTTTCATATGA 857
 723 TATGGAGTATCAGTTGCAAAGAACTTGATGAAAGTAGAAGTTTCCTCCTG 772

 858 GGATGCTGCAGATTCAATCTTCAAGATGGATGGTTACATCTCAAATGCAA 907
 773 TGACTCCTCTGGTTGTACTTTTGATATGGAGGGTTTCATATCCAATTCTA 822

 908 ATTATGTGGCAAAGAAGATTACAATGATTCTTTTCATAAATGATAGGCTT 957
 823 ACTATGTTGCTAAGAAGACTATATTGGTGCTTTTCATTAATGATAGATTG 872

 958 GTAGACTGTACTGCTTTGAAAAGAGCTATTGAATTTGTGTACTCTGCAAC 1007
 873 GTGGAATGCTCTGCCTTAAAAAGAGCCATTGAAATTGTTTATGCTGCAAC 922

1008 ATTGCCTCAAGCATCCAAACCTTTCATATACATGTCCATACATCTTCCAT 1057
 923 ATTGCCAAAAGCATCAAAACCTTTTGTCTACATGTCAATCAATTTGCCAC 972

1058 CAGAACACGTGGATGTTAATATACACCCAACCAAGAAAGAGGTTAGCCTT 1107
 973 GGGAACATGTTGATATCAATATTCACCCAACAAAGAAAGAGGTTAGCCTT 1022

1108 TTGAATCAAGAGCGTATTATTGAAACAATAAGAAATGCTATTGAGGAAAA 1157
1023 CTAAACCAGGAAATCATTATTGAGATGATACAGTCAGAGGTTGAAGTAAA 1072

1158 ACTGATGAATTCTAATACAACCAGGATATTCCAAACTCAGGCATTAAACT 1207
1073 ACTGAGAAACGCAAATGATACTAGGACGTTTCAAGAGCAGAAAGT..... 1117

1208 TATCAGGGATTGC..TCAAGCT........AACCCACAAAAG....GATA 1243
1118 ......GGAATACATTCAATCTACGTTAACATCTCAGAAAAGTGATTCTC 1161

1244 AGGTTTCTGAGGCCAGTATGGGTTCTGGAACAAAATCTCAAAAAATTCCT 1293
1162 CAGTTTCTCAG......AAGCCTTCTGGACAAAAGACACAGAAAGTTCCT 1205

1294 GTGAGCCAAATGGTCAGAACAGATCCACGCAATCCATCTGGAAGATTGCA 1343
1206 GTGAACAAAATGGTGAGAACAGATTCATCAGATCCAGCTGGAAGGTTACA 1255

1344 CACCTACTGGCACGGGCAATCTTCAAATCT...TGAAAAGAAATTTGATC 1390
1256 TGCCTTTTTGCAACCCAAGCCACAAAGTCTCCCTGACAAGGTTTCTAGTT 1305

1391 TTGTATCTGTAAGAAATGTTGTAAGATCAAGGAGAAACCAAAAAGATGCT 1440
1306 TGAGTGTAGTAAGGTCTTCTGTAAGGCAAAGAAGAAACCCAAAGGAAACT 1355

1441 GGTGATTTGTCAAGCCGTCATGAGCTCCTTGTGGAAATAGATTCTAGCTT 1490
1356 GCTGATCTTTCTAGTGTCCAGGAACTTATTGCTGGAGTTGACAGCTGCTG 1405
```

FIGURE 4B

```
1491 CCATCCTGGCCTTTTGGACATTGTCAAGAACTGCACATATGTTGGACTTG 1540
1406 CCATCCAGGTATGCTGGAGACTGTAAGGAATTGCACATATGTTGGAATGG 1455

1541 CCGATGAAGCCTTTGCTTTGATACAACACAATACCCGCTTATACCTTGTA 1590
1456 CAGATGATGTTTTTGCTTTAGTTCAGTATAACACCCATCTATATCTAGCA 1505

1591 AATGTGGTAAATATTAGTAAAGAACTTATGTACCAGCAAGCTTTGTGCCG 1640
1506 AATGTGGTGAATCTCAGCAAAGAGCTAATGTATCAGCAAACTCTTCGTCG 1555

1641 TTTTGGGAACTTCAATGCTATTCAGCTCAGTGAACCAGCTCCACTTCAGG 1690
1556 TTTTGCTCATTTTAACGCAATACAGCTTAGCGATCCAGCCCCTTTGTCAG 1605

1691 AGTTGCTGGTGATGGCACTGAAAGACGATGA.ATTGAT.....GAGTGAT 1734
1606 AGTTGATATTGTTGGCTCTGAAAGAGGAGGATCTAGATCCAGGAAATGAT 1655

1735 GAAAAGGATGATGAGAAACTGGAGATTGCAGAAGTAAACACTGAGATACT 1784
1656 ACAAAAGATGATCTGAAAGAAAGAATTGCTGAAATGAATACAGAACTCCT 1705

1785 AAAAGAAAATGCTGAGATGATTAATGAGTACTTTTCTATTCACATTGATC 1834
1706 CAAGGAAAAAGCAGAAATGTTAGAGGAGTATTTCAGCGTGCACATTGACT 1755

1835 AAGATGGCAAATTGACAAGACTTCCTGTTGTACTGGACCAGTACACCCCT 1884
1756 CCAGTGCAAATTTGTCAAGGCTTCCTGTGATACTCGACCAGTATACACCT 1805

1885 GATATGGACCGTCTTCCAGAATTTGTGTTGGCTTTAGGAAATGATGTTAC 1934
1806 GACATGGATCGTGTTCCTGAATTTTTACTATGCTTGGGAAATGATGTTGA 1855

1935 TTGGGATGACGAGAAAGAGTGCTTCAGAACAGTAGCTTCTGCTGTAGGAA 1984
1856 GTGGGAAGATGAGAAGAGTTGCTTTCAAGGAGTTTCTGCAGCTATTGGGA 1905

1985 ACTTCTATGCACTTCATCCCCCAATCCTTCCAAATCCATCTGGGAATGGC 2034
1906 ACTTTTACGCCATGCATCCTCCTCTTTTGCCAAACCCATCGGGTGACGGT 1955

2035 ATTCATTTATACA.......................AGAAAAATAGAGATTC 2063
1956 ATTCAGTTCTATAGTAAGAGAGGTGAGAGCTCTCAGGAAAAGTCAGATTT 2005

2064 AATGGCTGATGAACATGCTGAGAATGATCTAATATCAGATGAAAATGACG 2113
2006 AGAGGGTAACGTCGATATGGAGGACAATC..................... 2034

2114 TTGATCAAGAACTTCTTGCGGAAGCAGAAGCAGCATGGGCCCAACGTGAG 2163
2035 TTGACCAAGATCTTCTGTCAGATGCTGAAAACGCATGGGCACAACGTGAA 2084
```

FIGURE 4C

```
2164 TGGACCATTCAGCATGTCTTGTTTCCATCCATGCGACTTTTCCTCAAGCC 2213
     ||| | || || || || |||||||| || ||| || | ||| | |||||
2085 TGGTCAATCCAACACGTGTTGTTTCCGTCAATGAGATTGTTCTTGAAGCC 2134

2214 CCCGAAGTCAATGGCAACAGATGGAACGTTTGTGCAGGTTGCTTCCTTGG 2263
      ||     || |||||   || |||| || |||||   |||| || ||| | |
2135 ACCAGCTTCCATGGCTTCAAATGGACTTTTGTAAAGGTAGCATCCCTTG 2184

2264 AGAAACTCTACAAGATTTTTGAAAGGTGTTAGCTCATA 2301
     | || || |||||||| || ||| | || || || | |
2185 AAAAGCTGTACAAGATATTCGAACGATGCTAACTGAAA 2222
```

FIGURE 4D